DOCUMENT RESUME

ED 419 838 TM 028 414

AUTHOR	Brookhart, Susan M.
TITLE	Teaching about Grading and Communicating Assessment Results.
PUB DATE	1998-04-00
NOTE	16p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Diego, CA, April 14-16, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Communication (Thought Transfer); *Course Content; *Educational Assessment; Elementary Secondary Education; *Grading; Higher Education; Psychometrics; Standards; Teacher Competencies; *Teacher Education; Teaching Methods; *Test Results
IDENTIFIERS	Standards for Teacher Competence (AFT NCME NEA); Teacher Competencies

ABSTRACT

This paper discusses appropriate measurement content and instructional strategies for courses in classroom assessment in the areas of grading and communicating assessment results. Classroom teachers need to understand a wider range of assessments than many textbooks cover, and an aspiring teacher's classroom assessment practices need to be developed in concert with the instructional repertoire and classroom management skills. Important skills about communicating assessment results support Standards 5 and 6 of the "Standards for Teacher Competence in Educational Assessment" (1990). First, classroom assessment must be taught to aspiring teachers in relation to both instruction and classroom management, not simply as a decontextualized application of measurement principles. In the second place, the measurement content for classroom assessment courses has different emphases from the measurement content for introductory psychometrics courses. Third, the content of classroom assessment courses can best be taught by a mixture of direct instruction in the concepts and application examples and scenarios for classroom practice, simulation and discussion. Classroom assessment contributes to every other teaching function and helps create the classroom environment. (Contains 3 tables and 16 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 419 838

Teaching about Grading and Communicating Assessment Results

Susan M. Brookhart
School of Education
Duquesne University
Pittsburgh, PA 15282
412-396-5190
brookhart@duq3.cc.duq.edu

Paper presented as part of the Symposium:

Teaching Courses in Classroom Assessment

at the annual meeting of the

National Council on Measurement in Education

San Diego, CA

April, 1998

TM028414

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Susan Brookhart

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Teaching about Grading and Communicating Assessment Results

The purpose of this paper is to discuss appropriate measurement content and instructional strategies, for courses in classroom assessment, in the areas of grading and communicating assessment results. The content presented here is not meant to be an exhaustive course outline; rather, these examples are meant to illustrate some of the major differences in content between conventional educational measurement courses and classroom assessment courses.

The *Standards for Teacher Competence in Educational Assessment of Students* (1990) were developed jointly by NCME, AACTE, the AFT and the NEA. Standard #5 reads, "Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments." Standard #6 reads, "Teachers should be skilled at communicating assessment results to students, parents, other lay audiences, and other educators." The *Standards* considers both classroom assessment information and the results of external assessments under "assessment results." The *Principles for Fair Student Assessment Practices for Education in Canada* (1993) has similar concerns to the *Standards* that were developed in the United States. The Canadian document has two sections, "Classroom Assessments" and "Assessments Produced External to the Classroom." The Classroom Assessment section has standards for summarizing and interpreting results, which refers to "the procedures used to combine assessment results in the form of summary comments and grades which indicate both a student's level of performance and the valuing of that performance" (p. 10), and for reporting assessment findings. The External Assessment section has standards for interpreting assessment results and for informing students being assessed.

The rationale for a paper such as this, addressing some of the assessment competency needs for classroom practice, may be traced to the fact that many NCME members are the measurement or assessment specialists in the Schools, Colleges, or Departments of Education at their universities and are called upon to teach assessment courses for preservice or inservice teachers. This requires a different perspective on the measurement content than most measurement professionals received in their own education and training, which emphasized psychometrics for large-scale assessments. Absent any way to develop a perspective on the competencies required for classroom assessment, measurement experts sometimes just present simplified psychometric content in assessment courses for teachers. This is usually an unsatisfactory situation for both the professor and his or her students. The professor is left feeling like he or she trivialized important content. The students are left with information they can learn, but that does not directly apply to the classroom assessment they will be called upon to do. Students may mentally dismiss an instructor who does not demonstrate understanding of the classroom assessment context as lacking in credibility, thus minimizing their learning and retention of material from the class. NCME has been aware of this problem for some time (Nitko, 1991).

Given the importance of assessing well, it is crucial to attend to the quality of the assessment training given to pre-service teachers. One powerful way to do that is to give the measurement professionals who are called upon to teach assessment a perspective on what content is important for preparing teachers to do classroom assessment. Simplifying psychometrics is not the answer; principles for high-quality assessment, like validity and reliability, must be applied to the classroom context directly. Classroom teachers need to understand a wider range of assessments than many textbooks present and need to be offered methods that can be used within the constraints of classroom time and space and school district policies. An aspiring teacher's classroom assessment practices need to be developed in concert with his or her instructional repertoire and classroom management skills.

***Standards for Teacher Competence in Educational Assessment of Students* #5: Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.**

Communicating results is only as good as the quality of the message to be communicated. If classroom assessment information is of poor quality or incomplete, a teacher will not be able to effectively communicate information about student achievement. Other papers in this symposium address the kind of measurement knowledge and skills aspiring teachers need to develop or select, administer, and score classroom assessments. In addition, NCME has prepared some ITEMS modules that address individual classroom assessments (Arter & Spandel, 1992; Brookhart, 1993a; Stiggins, 1987, 1992).

At present, teachers must learn how to assign valid grades because the jobs for which they are being prepared require it. Teacher preparation in communicating the results of classroom assessment should take into account what schools do now and equip newly prepared teachers to help be part of needed change. Thus aspiring teachers need to know (a) how to assign letter grades or other report card symbols in ways that maximize validity and reliability and (b) how to communicate classroom assessment information in ways other than grades and how to advocate for change to these methods whenever that change would result in clearer communication of classroom assessment results. Many classroom assessment textbooks consider assigning grades as the only content under "communicating assessment results"; these texts may help instructors teach the former but they actively work against the latter, since they imply that grading is the only way to communicate information about classroom achievement.

Grading involves combining the results of assessments in ways that honor their intended weight in instruction and their informational value to the students. Norm-referenced weighting algorithms are usually not appropriate for objectives-driven instruction; simple criterion-referenced schemes (like averaging percents) may not work well, either. Combining test scores and rubric results in the same composite must be handled carefully. Despite the difficulties, preservice and inservice teachers must learn about grading because it is required in their professional practice. Other methods for communicating assessment results (exhibits, conferences, portfolios, and rubrics) should be taught and their use encouraged because of the limits of single letter grades.

Table 1 presents some content that aspiring teachers need to know in order to assign grades. Measurement professionals will note that much of this material is not different from material that might be taught in an introductory psychometrics course, but some of the emphases are different. An example of a major difference in emphasis for aspiring teachers compared with aspiring psychometricians is the conceptual treatment of validity, as compared with a more empirical treatment. Other content in Table 1 is different from what might be taught for psychometrics. Two examples of this are the combining of ordinal and interval measures and the choice of weighting methods for creating composite grades. Readers of this paper are urged to remember the purpose and context for these methods; the result in most grading applications is intended to be an ordinal-scale grade that reflects judgment of student achievement of instructional objectives. This is a very different target measure from most of the intended measures developed with psychometric methods.

Table 1
Examples of what aspiring teachers need to know about grading

Setting meaning for grades Understanding the relationship between model of instruction and mode of comparison Selecting the appropriate meaning for grades Identifying components for official assessment Developing compatible scoring scales for official assessments
Scaling component scores Understanding precision and rounding Choosing a scale appropriate to the assessment Writing rubrics and other scoring schemes Understanding level of measurement (especially Ordinal and Interval) Scoring failure and scoring failure to try
Combining component scores Knowing when to use mean and median Collapsing scales from more to less precision Transforming scales from interval to ordinal level Obtaining intended weights when forming a composite Matching weighting method to the intended meaning Reviewing borderline scores

Setting meaning for grades

Understanding the relationship between model of instruction and mode of comparison is important for deciding upon the appropriate grading model to use. It is not enough to teach students the measurement concepts that norm-referenced grading compares students to each other and criterion-referenced grading compares students with a standard, or even to add that self-referenced grading compares students with their own potential or progress. For classroom teaching, students must know that an objectives-driven model instruction implies that there should be some standards against which students may be measured. So students who learn instructional planning by writing unit goals and lesson objectives should understand that this fits with criterion-referenced grading. An older model of teaching, the transmission of information model, sometimes called "teaching as telling," can support a norm-referenced grading system. Students with varying backgrounds and interests in a topic will learn from lecture and text in ways that reflect their normally-distributed background experiences and interests.

Aspiring teachers need to be taught how to select the appropriate meaning (norm- or criterion-referenced) for grades (Frisbie & Waltman, 1992). Students need to discuss how these different models of instruction imply different approaches to grading. Most teacher education programs do not remain neutral on the subject of models of instruction, but rather advocate that

instruction should be based on goals, objectives, or achievement targets of some sort. Students learn how to implement this model in instructional planning courses. Thus, they should not be taught that the choice of grade meaning is a "choice" in the sense of a free pick. Teacher education programs that teach the use of instructional objectives should advocate the use of criterion-referenced grading and teach students several different ways to do that well.

A discussion of the differences between true criterion referencing and the simple calculation of percent-correct scores for an assortment of tests and assignments would be instructive in classroom assessment courses. It would make most sense to aspiring teachers if it were illustrated with lots of examples of real classroom assessments. Many curriculum materials have unit tests or worksheets that would make good examples for this purpose. Looking at these examples could also lead to a discussion of validity in the classroom context, highlighting that the achievement targets specified in instructional objectives must be clearly reflected in classroom assessments before percent-correct scores can be considered "criterion-referenced" in the sense of indicating what the pupil can do. It is then an additional step to broaden the construct from a single achievement target or unified set of them, as for one classroom assessment, to achievement on the entire set of instructional goals for a report period. The "construct" underlying reporting grades is then highlighted for discussion, and the degree to which a "criterion" can be specified at all would be open for discussion. This is a point at which aspiring teachers may develop some of the concepts they will need to argue as change agents in "reform" efforts in the schools where they will ultimately work.

Once the grading model is clarified, teacher education students need to also learn that not every assessment one does in a classroom should be used for summative grading purposes ("official" assessment, Airasian, 1994). For formative assessment during teaching, criterion-referenced and self-referenced student feedback are appropriate, the former for helping to create in the student's mind a concept of what quality ideas and performances look like, and the latter for helping the student gauge his or her progress toward quality (Harlen & James, 1997). So aspiring teachers need to learn how to provide self-referenced, descriptive feedback on assessments and also learn not to select these assessments for inclusion in summative grades.

Assessment results that should be included in composite grades should be criterion-referenced assessments that were administered after pupils had an opportunity to learn the knowledge or skills. These scores, and these alone, should comprise official assessment. Aspiring teachers should be warned that using "grades" as a tool for behavior management is not generally acceptable, but then they must be given alternative ways to insure that pupils complete their work and do their best. Teacher education, then, needs to coordinate the students' work in the area of assessment not only with their study of instructional planning but also with their study of classroom management.

Aspiring teachers not only need to know how to identify or develop appropriate official assessments that match their instructional intentions, they also need to learn how to develop compatible scoring scales for them. Percent-correct scores work well for tests or other "point"-based assignments that have at least 30 points and that are appropriately matched to instructional objectives. Rubrics work well for performance assessments, including written work, but do not mesh neatly with percents. Choosing a scale appropriate to the assessment is a topic not covered often enough in classroom assessment courses.

Scaling component scores

Aspiring teachers need to know enough about precision and rounding that they do justice to the type of measure they have. They need to learn that they can transform scores in the direction of less precision (e.g., from percents to letter grades or to rubrics) but they cannot move in the opposite direction. They need to coordinate rubrics for different assignments that will ultimately be

combined in such a way that the quality levels are compatible. It is important to teach the quantitative reasoning behind these principles so as various problems of application arise, teachers can solve them. Most aspiring teachers, whether they have encountered precision, rounding and mapping one scale onto another in a mathematics class or not, will not automatically use these concepts in their working repertoire. Classroom assessment instructors should review these concepts and show students how to apply them specifically to grading.

Writing rubrics and other scoring schemes require special verbal as well as quantitative skills. Choosing the numerical levels for a rubric or deciding how many points (and therefore what weight) to give to various components of a scoring scheme must be done with an eye to validity, in this case most importantly a match of scoring emphasis with instructional intent. But beyond that, the verbal descriptions that go with rubric levels and the directions for use in other kinds of point scoring schemes require clear communication of the concepts or performances assessed, that is, clear descriptions of what high quality work looks like. It takes practice to write these well. Lacking clear writing, neither teachers nor students will be able to use the rubrics reliably, because it will not be clear what each level means. Validity too will suffer, since it is hard for something that is imprecise and poorly expressed to represent instructional intent. Here is another clear link between teaching aspiring teachers about assessment and teaching them about instruction. As Judy Arter writes, "The single biggest issue facing teachers as they design assessments has nothing to do with assessment per se, but with having a clear understanding of the learning targets they should have for students" (Arter, personal communication, 1/28/98).

Understanding level of measurement (especially ordinal and interval levels) is more important to teacher education now than ever. The deserved popularity of rubrics, most of which use ordinal level scales, has caused some consternation. In the eight grade in a school district this author works with, teachers were faced with the task of combining percent-correct scores from conventional Language Arts tests and writing performances scored on a 4-point rubric into 5 levels for report card grades (A,B,C,D,F). Several of them did not have the quantitative reasoning background to understand why or how scale conversions could be made, and it had not occurred to any one of the several people who adopted the 4-point writing rubric that it would not be very helpful for assigning five levels of grade. This is a more complicated problem to solve after the fact than to solve at the design stage, when it would be appropriate to choose rubrics and construct decision rules.

The quantitative concepts behind level of measurement, precision, and scaling may seem foreign to some teacher education students, many of whom will have had a rote approach to mathematics in their own backgrounds. But these concepts offer some rich, interesting, potentially even "fun" classroom activities in the classroom assessment classes. Students or groups of students can work with scenarios, either real like the one just described or hypothetical, devise solutions and discuss them, and try applying them to samples of student work. In the author's experience, teacher education students see real value in simulations of real classroom tasks. The instructor's contribution is to facilitate the discussion and to make explicit the concepts about level of measurement, precision, and scaling as they arise, making suggestions for improvement if they are incorrectly applied and articulating a justification when the concepts are rightly applied if the students do not offer one themselves.

Scoring failure and scoring failure to try are issues that can generate emotional responses from teacher education students. Information for quantitative reasoning and information about instruction and assessment, in concert, will give aspiring teachers the tools they need to solve the failure and failure to try issue, one instance at a time. What does it mean to give, say, a "50" to unacceptable quality work (an F) and a zero for failure to hand in work, on the same scale? Should rubrics use the same level, typically "1," for unacceptable and missing, as many do? The practice of assigning a zero to missing work can be explored via scenario in classes. Groups of students can be assigned to work out various good and bad solutions to different versions of the

problem, including scenarios about students who forgot, students who were truly resistant, students who had been counseled about missing work before, students who were using learning contracts: (a) assigning a zero and calculating a mean final grade, (b) using the median method with grades on each assignment instead of percent scores (which will precipitate, perhaps, a discussion of how much precision of information is available in a classroom test and whether or not the implied hundred-point continuum of percents accurately captures that), (c) give no grade for the missing assignment and calculate the final grade on the basis of other assignments, (d) give the missing assignment a 50 (the bottom of an F range that would be the same size as the other intervals) and calculate a mean final grade, (e) counsel the student about work habits or keep him or her after school to do the assignment, and (f) make the student do a make-up assignment in class. The criterion for judging whether a solution is "good" or "bad" will be the extent to which the grade communicates clear information about the student's achievement of the instructional intent for the reporting period, and should take the discussion back to validity.

Longer-term solutions, like reform of a school district policy that brooks no Incompletes (unlike the college course the students themselves will be taking) or, even more radical, reform of a grading system that requires grades for all students at the same time, can be discussed, too, so that students see that the "missing data" problem in grading is in some respects an artifact of policies and assumptions about the conduct of education more than a measurement problem. The point is not to teach that there is a good solution to the problem as it stands in schools today, but rather to develop the measurement and instructional and management skills, in concert, for approaching the problem.

Combining component scores

As the rubric/percent discussion and the missing data problem above both imply, knowing when to use mean and median is an important measurement tool for those who must calculate component grades. The median is a good measure of central tendency to use with the ordinal level data or, more commonly, the mix of ordinal and interval level data that comprise most official assessment scores for grading. Even the scales that look like interval level scales, for example number right or percent correct, often appear to have more precision than they actually do. A more appropriate match to the kind of information is often a letter grade; a set of recorded letter grades can be conveniently and defensibly summarized with a median. This is a method not often used, and the author wonders why, since it seems to fit "classroom reality" (Airasian, 1991) so much better than many grading methods that are used. Perhaps it is simply that most aspiring teachers were never given this tool to put in their repertoire. Information about collapsing scales from more to less precision and transforming scales from interval to ordinal level can be taught with the instruction about level of measurement, since these are practical applications (and would make good class exercises) that will demonstrate to aspiring teachers the reason for learning the material.

Obtaining intended weights when forming a composite grade is an important issue that goes directly to validity. The composite grade needs to match the instructional intent of a reporting period considered as a whole. Composites not weighted in a way that comports with the instructional intent of the reporting period are, arguably, not valid for their intended purpose and not fair to students. Aspiring teachers need to learn how to match the weighting method to the intended grade meaning (Oosterhof, 1987). When composite grades are calculated as means, the weight of components is affected by their variability when grades are intended to be norm-referenced and by maximum possible points when grades are intended to be criterion-referenced. Aspiring teachers should learn at least how to do maximum-possible-points weighting. If the teacher education program teaches an instructional-objectives method of teaching and therefore advocates criterion-referenced grading, it would be wiser to spend available time teaching how to weight when using the mean and the median for grading than to take a lot of time teaching algorithms for weighting by variability. Weighting on the basis of variability should be explained

conceptually, however, since teachers will need to check gradebook programs they may use to see which method is the default and whether the method they would choose is an option.

Thus far, this paper has considered mainly quantitative concepts for classroom assessment courses that cover grading. Another area for study is one with which measurement instructors may not have as much experience, and that is teacher professional judgment. Even the most mechanically computed grades are not judgment-free, since a teacher plans what instruction and assessments to use for reasons that involve educational judgment. Adjusting what components go into the official assessment for grades according to individual student needs and/or adjusting individual component assessments also require judgment. Applying rubrics reliably involves professional judgment and will be discussed below.

Reviewing borderline scores is another area that requires professional judgment. The nature of that judgment, when, why, and how to review borderline scores, should be the focus of at least some study. Many teachers find it comfortable to review "just under" borderline scores and adjust them upward but would not think of doing the opposite (Brookhart, 1993b). Aspiring teachers should learn the concept of measurement error and learn to accept that review of borderline scores may be justified. They also need to conceptualize this review in validity terms, so that the additional information they consider in a borderline review comports with the information the grade is meant to convey, the instructional intent of a reporting period considered as a whole. Thus additional information about achievement of that instructional content is more relevant for a borderline review than additional information about a student's level of effort.

All of these grading concepts may be taught with a mixture of direct instruction and active application. Group work designing hypothetical grading plans, in the author's experience, is less helpful than work on scenarios and real work samples. Absent a particular "word problem" to work on, aspiring teachers sometimes design things that are too general to give them practice working with the concepts just described. Asking "why" and "what else could you have done" are important for application work. Students who are asked to reflect are also being asked to put their ideas into words, and that will help turn their classroom assessment learnings into knowledge they will be able to remember and skills they will be able to use.

***Standards for Teacher Competence in Educational Assessment of Students* #6: Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.**

Methods other than grades for communicating classroom assessment results (exhibits, conferences, portfolios, and rubrics) apply under this standard and have been advocated in the previous section. These methods of communicating information about student achievement and progress require that aspiring teachers have good written, oral, and interpersonal communication skills.

Course content that teacher education might address, in addition to grading, to equip aspiring teachers to communicate classroom assessment results and information about student achievement are listed in Table 2. This list contains examples and is not meant to be exhaustive; nevertheless, note how much of the content is not what would be emphasized in an introductory psychometrics course.

The measurement concept behind most of the items on the list is validity. A measurement instructor who teaches aspiring teachers should be prepared to teach students how to do these things and to argue for how careful attention to these tasks would enhance validity. An understanding of the concept of a construct and a working repertoire of examples of "constructs" that are common in classrooms would help in instruction. Thus, for example, instead of explaining constructs as the shared variance among a group of measures of a latent variable, it

would be helpful to explain constructs as performance on the "achievement targets" or objectives of classroom instruction, or the interests and attitudes of students, and so on. The author has found that preservice teachers and inservice teachers both find Stiggins' (1992) "achievement target" metaphor very helpful.

Table 2

Examples of what aspiring teachers need to know about communicating classroom assessment results in ways other than grades

Portfolios

Articulating achievement targets (objectives)
Articulating the qualities of good work and helping students learn to recognize these in their own work
Talking with students about work
Listening to students talk about their about work
Teaching students how to reflect on the quality of their work

Conferences

Parent--teacher
Student--teacher
Student--parent--teacher
Interpersonal communication about academic work
Articulating the qualities of good work and/or expectations for student learning and behavior
Communicating the results of comparing one student's work against these criteria
Listening to student and parent responses

Exhibits

Articulating the qualities of good work and helping students learn to recognize these in their own work
Selecting examples to exhibit and being able to articulate the reason for the selection

Rubrics

Articulating the qualities of good work in a descriptive continuum
Disentangling judgment and description, then doing both well
Observation and judgment skills regarding students working and the products of their work
Identifying when, and knowing how, to use different kinds of rubrics (holistic or analytic, generalized or task specific)

Portfolios

Portfolios are widely used in classrooms nowadays, so aspiring teachers should learn how to use them. Portfolios, however, are like meat loaf; different educators have different recipes for them. Only some of the purposes and uses of portfolios emphasize assessment; some kinds of portfolios have largely instructional functions. Some kinds of portfolios have no summative assessment purposes at all and are purely formative, for example writing portfolios in which pupils reflect on their own work and try to improve it. So one of the tasks a classroom assessment instructor has is to identify the different kinds and purposes of portfolios and to link these with both instructional concepts and measurement concepts (Nitko, 1996). There are all kinds of ways that formative assessment can take place within and through the use of portfolios. Some of this assessment is criterion-referenced, as when writing rubrics are applied to pupils' work. A lot of it is self-referenced. Much of the power of portfolio assessment, from a learning theory point of view, is in the student's role as assessor of his or her own work (Arter & Spandel, 1992). Classroom assessment courses should address all these functions because all of them are relevant to the effective use of assessment in classrooms.

Some of the knowledge and skills that are absolutely crucial to the valid and effective use of portfolios for assessment purposes requires teaching things that measurement instructors may be more used to viewing as topics for classes in instructional planning, instructional methods, or even English and communication classes. Measurement principles must be integrated with instructional principles and classroom management principles. One way for a measurement professor to do this is to plan a panel discussion with instructional and management professors. Another way is through assigning readings that cross these boundaries. Yet without instruction in, and practice with, these things, aspiring teachers will not be able to use portfolios well, even for assessment purposes. The reliability and validity of a measure suffers when the students being assessed are not clear about what is being asked of them.

So teachers need to practice articulating achievement targets in terms that students can understand (Stiggins, 1997) and working to understand the achievement target completely themselves. Teachers who cannot write well, or at least recognize good writing when they see it, will not be able to assess pupils' writing with portfolios. Articulating the qualities of good work and helping students learn to recognize these qualities in their own work, while necessary for all good instruction, bleeds into assessment when portfolios are the assessment vehicle. Similarly, talking with students about work, listening to students talk about their about work, and teaching students how to reflect on the quality of their work may seem to belong more properly in a classroom management class or even a communication class, but these tasks need to be done well in order to support the validity of an assessment of student achievement based on a portfolio. These may be areas that an assessment instructor never expected to have to teach. Working with other faculty members or local school teachers may be helpful.

Conferences

Conferences can be another means of communicating achievement information. They hold special promise because the communication is interactive and because there is the potential for selecting different pieces of information or even different themes to discuss for different pupils' conferences. Conferences are time-consuming. Important for the classroom assessment instructor, conferences about achievement must be based on a presentation and discussion of evidence. Gathering, interpreting, and presenting that assessment information are skills that aspiring teachers should have. Practice in conference simulations in class would be helpful. At the least, preservice teachers should practice gathering and interpreting the evidence for a conference, even if there is not time to role-play conferences in class.

There are at least three kinds of conferences that classroom assessment courses should consider (Stiggins, 1997): parent--teacher, student--teacher, and student--parent--teacher. Each has its own dynamics. Interpersonal communication about academic work is a skill that has been relatively neglected in teacher education, in both assessment and instruction courses. Preservice teachers would benefit from practice at the kind of language and approaches that are helpful when sharing information about student achievement. Actively listening to students' and parents' responses requires practice, too. As with all classroom assessment, articulating the qualities of good work and/or expectations for student learning and behavior is crucial. But it takes on a special urgency when these criteria must be articulated to parents in person. Communicating the results of comparing one student's work against these criteria and listening to student and parent responses also require practice.

Exhibits

Exhibits can be a good way to communicate information about student achievement to a community audience. The sports and fine arts departments in schools have long had athletic events, concerts and plays for parents and interested community members to attend. These events at least tacitly communicated some information about "what students can do" to those who were watching. Exhibits that are expressly for the purpose of communicating what students can do in academic tasks are increasing in popularity. The author works with a district that has a portfolio fair in several of its grades. Parents come and hear students talk about the work in their portfolios. The author also once visited a second grade teacher whose students "publish" books, which are then read to parents at a tea. Again, articulating the qualities of good work, helping one's pupils learn to recognize these in their own work, and helping pupils select the examples to exhibit and articulate the reason for the selection are assessment related skills that preservice teachers need to be taught.

Rubrics

Writing rubrics well is a difficult task that is, in the author's opinion, worth the effort. Articulating the qualities of good work on a descriptive continuum is a skill with which some aspiring teachers will struggle. Many will want to use judgment words ("excellent, good, fair, poor") as the levels of achievement. The critical skill of disentangling judgment from description, then doing both well, is hard to teach. Classroom assessment instructors will need to assign aspiring teachers to write rubrics based on their expectations for good work and their conception of what hitting the achievement target would look like, and what near misses and stray shots would look like, in words that pupils could understand. Instructors will find that as for most writing assignments, editing and revising will be necessary. Clarity of writing is important in rubrics not only for its own sake, but for validity (since the "top category" will describe what the students learn is the instructional intent) and reliability (since if performance descriptions at the various levels are not clear, they cannot be reliably applied to pupils' work). Once clear rubrics are written, yet another set of skills is required: observation and judgment of pupils at work and observation and judgment of the products of pupils' work. This skill of rater reliability may be taught in classroom assessment courses in a similar fashion to the way it is taught in much rater training, using work samples to categorize and discussing why each is scored as it is.

Another measurement task that intersects with instructional planning skills is identifying when, and knowing how, to use different kinds of rubrics (holistic or analytic, generalized or task specific). Aspiring teachers must learn the purposes and uses of each of these. This author's opinion is that classroom assessment instructors ought to advocate the use of generalized rubrics whenever possible, making students aware that they are more difficult to apply reliably and giving them strategies for developing their skills at reliable scoring. The reason for this opinion is that it is in generalized rubrics that the "achievement target" or conception of good work is expressed. And this is the purpose of most education, not that the student can "do" an individual lesson but

that he or she learns some more general skill that the lesson exemplifies. So, for example, a math teacher may wish for a student to learn that a student has solved a word problem "well" when he or she has completely and correctly interpreted the problem elements, generated a strategy that will lead to the solution, and correctly implemented the solution. Such language may form the performance description of a generalized math problem-solving rubric. A task specific rubric for one problem would have the particulars of the problem within it. It would be harder for pupils to see the general elements of good problem solving. It is also not possible to share task-specific scoring rubrics with pupils as part of instruction, while generalized rubrics should be shared with pupils.

Communicating Standardized Test Results

Another aspect of standard #6 is that teacher education students should learn to communicate the results of standardized achievement tests to parents, students, and other educators. Communicating results is only as good as the quality of the message to be communicated. If faulty or incomplete conclusions are drawn because of misunderstanding of assessment information, a teacher will not be able to effectively communicate information about student achievement. In courses on classroom assessment, aspiring teachers need to learn the skills a classroom teacher needs to understand and use standardized test results for classroom instruction and to interpret standardized test results to parents.

The author reviews classroom assessment textbooks for a publisher and has seen in book prospectuses arguments both for and against including information about standardized testing in classroom assessment textbooks. Since the classroom teacher is likely to be the first one called if parents have a question, and since some information from standardized tests results can be used in classroom instructional decisions, it seems that basic information to interpret individual scores is important for aspiring teachers to learn. Information about aggregated scores and sampling is not as relevant to classroom teaching. Assessment instructors should not expect standardized test content to be primary information for teachers, nor should it consume a large portion of a classroom assessment course. The emphasis in classroom assessment courses should be assessing student achievement of classroom instruction.

Course content that an interpreter of scores needs will have different emphases than course content that a test developer needs. Table 3 contains examples of some of the content that should support aspiring teachers' work toward communicating results of standardized assessments. Preservice teachers should study the definitions of percentile ranks, stanines, and scaled scores, and know the uses for each. They do not need to know how to compute the various kinds of scores. But for many measurement instructors, their own concepts of the scores and their meanings were developed by learning how to compute them. It is important for measurement instructors to develop other ways of communicating these concepts to students.

One strategy that has worked for the author for teaching the meaning of scores without teaching their computation is to start with the score, translate it into words (which of course uses quantitative concepts), and then ask what such a score might mean for the child and for the teacher. For example, aspiring teachers should learn that a percentile rank of 60 means that the student scored as well as, or better than, 60 percent of students in a norm group. What does that mean for the student? The aspiring teachers can then discuss that it depends who is in the norm group, what kind of test, what purpose the score would be used for, and so on.

Table 3
Examples of what aspiring teachers need to know about standardized assessments

Scores and score meaning Status measures: percentile ranks, stanines Growth measures: standard scores, grade equivalents and age equivalents Norm groups and their implications for norm-referenced scores
Uses and misuses of information Interpreting confidence bands Generalizing and reasoning to the construct and not beyond Age appropriateness Difference between individual score reliability and decision accuracy Difference between grade equivalent score and grade-level instructional objectives

Scores and score meaning

The concepts of a norm group and of norm-referencing, and the difference between status and growth measures, are basic information for classroom teachers. The more common scores, and the ones in most general public use, should be stressed. Status measures most often used in schools are percentile ranks and stanines. Growth measures most often used in schools are standard scores and grade equivalents. Actually, grade equivalents are used more commonly than standard scores, but the classroom assessment instructor can advocate for better use of standard scores and less emphasis on grade equivalents. The difference between "expected performance for a student in that grade," a legitimate interpretation of grade equivalents, and "performance expected from a student in that grade," implying grade-level objectives and thus a misinterpretation of grade equivalents, is too fine a hair for many people to split. At the present time, the difference between grade equivalent score and grade-level instructional objectives is not well understood by the lay public and not well explained to them by classroom teachers.

Uses and misuses of information

Concepts that are important to the interpretation of individual pupils' standardized test results include interpreting confidence bands. Teach students how to do that, not how to calculate the bands. Another important idea for interpreting standardized tests is generalizing and reasoning to the construct and not beyond. Students should learn to ask what a standardized test is designed to measure and then make inferences and communicate results accordingly. Age appropriateness of tests, including at what age school districts may reasonably begin a standardized testing program, is a concept classroom teachers may enjoy discussing. The difference between individual score reliability and decision accuracy is another point classroom teachers need to understand. A child's score may be very reliable, but the use of that score to make a particular decision about the child's educational placement may be less reliable. Students should learn to once again ask whether what the test was designed to measure is the relevant input for the decision in question and what other information is important for the decision. Standardized tests should be portrayed as tools for providing information, along with other achievement and work habits information and teacher judgment.

Summary

This paper has presented suggestions for the kind of content and instruction that classroom assessment courses should contain regarding (a) communicating achievement results by assigning grades, (b) communicating assessment results in ways other than grades, and (c) communicating the results of standardized testing. These important skills about communicating assessment results support Standards #5 and #6 of the *Standards for Teacher Competence in Educational Assessment of Students* (1990). The presentation has been organized by topic area.

Three themes cross all the topic areas. First, classroom assessment must be taught to aspiring teachers in relation to both instruction and classroom management, not simply as a decontextualized application of measurement principles. A measurement instructor without much training in recent work on instructional strategies or classroom management may wish to work with colleagues or guest teachers.

Second, the measurement content for classroom assessment courses has different emphases from the measurement content for introductory psychometrics courses. This paper has given some examples of what the author feels are some of the more salient differences in emphases. The point of view expressed is based on the author's work with preservice teachers, inservice teachers, and school administrators, on her research about classroom assessment, and on her own experience as a classroom teacher and teacher educator. There is room in this discussion for other perspectives, and in any case the content selected for this paper is not meant to be an exhaustive content outline for a classroom assessment course.

Third, the content of classroom assessment courses can best be taught by a mixture of direct instruction in the concepts (lecture, text), and application examples and scenarios for classroom practice, simulation, and discussion. There are at least three reasons for this: the general principle that practice with examples of any concept aids learning; the fact that many of the assessment competencies classroom teachers need are skills; and the particular case in teacher education where students have a well-documented interest in practical application to children's learning (Brookhart & Freeman, 1992).

Classroom assessment is a vitally important teaching function. It contributes to every other teaching function. Assessment helps create the classroom environment (Stiggins, 1997). It is in the best interests of the children who will be their students' pupils that NCME members deliver credible, useful, and sound instruction in classroom assessment content and skills in the courses they teach.

References

Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues & Practice, 10* (1), 13-16, 26.

Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.

Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment (ITEMS module). *Educational Measurement: Issues & Practice, 12*(1), 36-44.

Brookhart, S. M. (1993a). Assessing student achievement with term papers and written reports. (ITEMS module). *Educational Measurement: Issues & Practice, 11*(1), 40-47.

Brookhart, S. M. (1993b). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30,* 123-142.

Brookhart, S. M., & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research, 62,* 37-60.

Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan (ITEMS module). *Educational Measurement: Issues & Practice, 11*(3), 35-42.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4,* 365-379.

Nitko, A. J. (1991). Editorial: What are we teaching teachers about assessment and why? *Educational Measurement: Issues & Practice, 10* (1), 2.

Nitko, A. J. (1996). *Educational assessment of students* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Oosterhof, A. C. (1987). Obtaining intended weights when combining students' scores (ITEMS module). *Educational Measurement: Issues & Practice, 6* (1), 29-37.

Principles for fair student assessment practices for education in Canada. (1993). Edmonton, Alberta: Joint Advisory Committee.

Standards for teacher competence in educational assessment of students (1990). American Federation of Teachers, National Council on Measurement in Education, and the National Education Association.

Stiggins, R. J. (1987). Design and development of performance assessments (ITEMS module). *Educational Measurement: Issues & Practice, 6*(3), 33-42.

Stiggins, R. J. (1992). High quality classroom assessment: What does it really mean? (ITEMS module). *Educational Measurement: Issues & Practice, 11*(2), 35-39.

Stiggins, R. J. (1997). *Student-centered classroom assessment* (2nd ed.). Upper Saddle River, NJ: Merrill.